DOCUMENT RESUME

ED 027 447 AC 003 051

By-Merrill, Irving R.; Yaryan, Ruby B.

Broadcast TV as an Aid to Continuing Education. Terminal Progress Report.

California Univ., San Francisco. Medical Center.
Spons Agency-National Heart Inst., Bethesda, Md.
Report No-PHS Grant-HE-09564
Pub Date 30 Jun 68
Note-42p.
EDRS Price MF-$0.25 HC-$2.20
Descriptors-Achievement, *Adult Learning, *Attendance, Films, Physicians, *Professional Continuing Education, Programed Instruction, Reinforcement, Retention, *Scheduling, Statistical Data, *Televised Instruction, Time Factors (Learning)

The effect of massed versus distributed television presentations on attendance and learning in a voluntary situation were compared in a field experiment involving 114 physicians in general practice. The factorial design was based on four experimental groups, each located in a different community. Physicians viewed videotape presentations in a local hospital over a period of four weeks, with three one hour sessions (distributed viewing) and a single three hour session (massed viewing). On the fifth week they completed a learning achievement test over televised instruction. The results indicated that attendance was significantly greater under massed than under distributed viewing, especially when the sessions took place in the evening rather than the morning. The presentations produced gains in learning achievement, and there was no statistically significant difference in learning achievement between massed and distributed viewing. (author/nl)

BROADCAST TV AS AN AID TO CONTINUING EDUCATION

Report Prepared by:

Irving R. Merrill and Ruby B. Yaryan

University of California San Francisco Medical Center

BROADCAST TV AS AN AID TO CONTINUING EDUCATION

Terminal Progress Report
PHS Grant No.: HE 09564

Report Prepared by:

Irving R. Merrill and Ruby B. Yaryan

Principal Investigator: Irving R. Merrill

Co-Investigator: Seymour M. Farber

June 30, 1968

This Investigation was Supported by a Grant from
The National Heart Institute

Grantee:

University of California San Francisco Medical Center

TABLE OF CONTENTS

I. SUMMARY AND ACKNOWLEDGEMENTS 1

II. RETENTION AND ENJOYMENT OF PROGRAMMED INSTRUCTION: THE EFFECTS OF EFFORT AND REINFORCEMENT 5

III. EFFECTIVENESS OF MOTION PICTURES AT DIFFERENT STAGES OF LEARNING HISTORY-TAKING 17

IV. TELEVISION RECORDINGS FOR PRACTICING PHYSICIANS: THE EFFECT OF PROGRAM LENGTH ON ATTENDANCE AND LEARNING 27

APPENDIX 38

REFERENCES 39

* * *

I. SUMMARY AND ACKNOWLEDGEMENTS

Summary

This Terminal Progress Report describes the research conducted between May 1, 1965, and December 31, 1967, under a grant from the National Heart Institute of the National Institutes of Health. The research program is described in Application HE 09564-03, which was submitted October 13, 1966.

Research in the use of television as an aid to continuing education should be conducted in such a way that the findings are significant and of broad relevance. In an attempt to achieve these goals, the project included a Co-Investigator who is widely known as a medical educator directly involved in all areas of continuing health sciences education. It also included a distinguished Advisory Committee which was drawn from the fields of medicine, dentistry, nursing, and health sciences research. Committee members participated on a personal basis, not as representatives of an organization. The assistance of these persons is acknowledged elsewhere, and their names, along with those of the Project Staff, are listed in the Appendix.

Summary Statement of Progress

In brief, the objective of this research has been to initiate some investigation in three major areas which affect the efficiency of television in continuing medical education. Each of these areas affects what a physician learns and remembers from viewing a television presentation. One area also affects the physician's decision to view, an important consideration in that almost all continuing medical education is based upon voluntary physician participation.

The first area involves factors which are basic to the psychology of all learning. The study measured the effect of effort and reinforcement on initial learning and on learning subsequently retained by medical personnel. It also measured the effect of effort and reinforcement on the degree of enjoyment which was derived from the learning tasks. The subject matter to be learned dealt with anaphylaxis and classification of allergies, and the means of experimental presentation was the programmed text.

The second area involves factors related to visual information, i.e., graphic communication. The study measured the effectiveness of eight types of visual presentation at different stages of the learning process. The subject matter dealt with communication techniques in patient history taking, and the means of experimental presentation was the motion picture film.

The third area involves practical adjustments which the medical educator can make in the time at which instruction is made available. The study measured the effect of length and frequency of presentations on the attendance and the learning of general practitioners in the "real world" of emergency calls, conflicting interests, and even physician fatigue. The subject matter dealt with pulmonary embolism, chromosome studies, myocardial infarction, cirrhosis, low-back pain, and multiple sclerosis, and the means of presentation was the playback of recordings made on videotape.

These three studies were completed on schedule, and as a result, the stated aims of the project were achieved. Each of the following chapters describes a particular study in detail, including a statement of the specific problem, a description of the methodology followed, and a report and discussion of the findings obtained.

Significant Results

There are about 20 findings reported in the chapters to follow which are considered significant. Five of this number may be considered to have particular interest for the reasons indicated:

1. Learning achievement is greatest from a medical motion picture if the information is abstracted (as in a line drawing) when presented at the early stages of learning and if the information is presented realistically (as in a photograph) at the later stages of learning.

--This finding is extremely significant because it may generalize to television and other media of graphic communication, regardless of the type of subject matter, i.e., medicine, chemistry, etc.

--This is the first reported test of the hypothesis.

--Current theory and practice in the preparation of visual messages generally fail to consider the stage of learning of the intended viewing audience.

2. Significant gains in learning resulted when groups of general practitioners viewed the playback of videotape recordings in a hospital conference room, as measured by a 120-item objective test.

--This is the first reported study of the effectiveness of videotape for a voluntary audience of physicians in active practice.

3. Physician attendance was significantly greater for a single

three-hour (massed viewing) session than for three weekly one-hour (distributed viewing) sessions, when held in the evening.

--This information has practical significance for the medical educator who must use limited funds for the benefit of the maximum number of physicians.

4. Massed viewing resulted in no significant reduction of learning achievement when compared with distributed viewing of the identical television recordings.

--This information is of practical significance to the medical educator who is interested in efficient instruction as well as in a large number of registrants.

5. When retested after a month, medical students remembered more from programmed instruction if they had worked through the text as the authors intended, i. e., manipulation of the psychological variables of effort and reinforcement either interfered with student memory or failed to improve it.

--This is a significant negative result which tends to eliminate two theoretically plausible means of improving physician memory of newly achieved medical knowledge.

List of Publications

An article has been prepared to cover the research conducted in each of the three major areas and has been submitted for publication. The current status of these articles is as follows:

Merrill, I. R., Yaryan, Ruby B., & Musser, T. S. Retention and enjoyment of programmed instruction: the effects of effort and reinforcement. Unpublished manuscript, University of California San Francisco Medical Center, 1968.

Merrill, I. R., Yaryan, Ruby B., Carbone, J. V., Musser, T. S., & Vandervoort, H. E. Effectiveness of motion pictures at different stages of learning history-taking. Paper scheduled for presentation at the Seventh Annual Conference on Research in Medical Education, Association of American Medical Colleges, Houston, November 1, 1968.

Merrill, I. R., Yaryan, Ruby B., Musser, T. S., & Ury, H. K. Television recordings for practicing physicians: the effect of program length on attendance and learning. Unpublished manuscript, University of California San Francisco Medical Center, 1968.

Acknowledgements

Grateful acknowledgements are due Seymour M. Farber, M.D., who served as a Co-Investigator of the project, Roger H. L. Wilson, M.D., Donald L. S. Kimbrough, M.D., Mrs. Audrey Peterson, and the other staff of Continuing Education in Medicine and the Health Sciences for their assistance with the research, especially the study directly related to physicians in general practice.

Special thanks are due the members of the Research Advisory Committee for their encouragement and wise council during the project. The loyalty and assistance of the Project Staff listed in the Appendix and other staff members of the Communications Office for Research and Teaching deserve special mention, for without their help the project could not have been completed on schedule.

The contributions of J. V. Carbone, M.D., and H. E. Vandervoort, M.D., in providing the subject matter on the history-taking and their assistance in the study warrants special recognition. Special thanks are due Dr. Ellen Brown and Dr. C. T. Carman who helped with arrangements for the main experiment and Prof. R. Warwick, Guy's Hospital Medical School, who helped with arrangements for the independent study of the materials on history-taking.

The statistical consultation provided by Hans Ury, M.A., was greatly appreciated. Thanks are due the numerous assistant experimenters and projectionists, whose assistance in the conduct of the research was invaluable.

Grateful acknowledgement is due Leon Summit, Editor of Pfizer Spectrum, who supplied copies of the program Allergy and Hypersensitivity used in the investigation of the factors of effort and reinforcement, and the Department of Microbiology, whose cooperation in the experiment was necessary for its completion.

Special thanks are due the California Academy of General Practice and the 138 practicing physicians who were so cooperative during the videotape viewing study, as well as John Lehotsky, representing Ward/Davis Associates for furnishing videotape recorders and monitors.

Thanks are due the many members of the faculty and staff of the University of California Schools of Medicine and Dentistry for their assistance with the research, and the Medical and Dental School Classes of 1968 and 1969 for their participation in two of the projects. Finally, it is a pleasure to acknowledge the splendid cooperation of the faculty and the enthusiastic participation as subjects by 131 students at Guy's Hospital Medical School, University of London.

II. RETENTION AND ENJOYMENT OF PROGRAMMED INSTRUCTION: THE EFFECTS OF EFFORT AND REINFORCEMENT

Abstract

The effects of two degrees of effort and four schedules of reinforcement on initial learning, enjoyment, and learning retained after one month were compared for eight groups of medical and dental students. The instructional material was a programmed medical text. The purpose of the main experiment was to test the dissonance hypotheses that (a) the requirement of a high degree of effort in working through the program produces greater enjoyment and greater retention of learning than does a low degree of effort and (b) initial learning and retention are not affected by the difference between fixed and variable reinforcement at the same ratio. In a pilot experiment the instructional effectiveness of the program was measured against a control group. The results on enjoyment and retention were inconclusive, but the prediction regarding fixed and variable ratio reinforcement was supported. The medical text was not very helpful to dental students, but it produced an impressive increase in learning for the medical students, whose reaction to this type of instruction was favorable provided it was not used as a substitute for teaching-contact with the faculty.

* * *

In a previous experiment to improve retention of learning (1), it was found that nursing students who were required to expend considerable effort while learning enjoyed the learning situation more than those who were required to expend less effort while learning the same material. It was also found that the high effort group tended to remember more after 40 and 59 days than did the low effort group, although this difference failed to reach the customary .05 level of statistical significance.

Because these results are not predicted by traditional reinforcement theory, it was considered desirable to replicate the experiment using a different teaching device and different students, increasing the effort required, and shortening the retention interval to approximately a month.

It was known that the faculty in a basic science department desired an objective evaluation of the suitability of a programmed text for use by medical and dental students. The text was popular and highly regarded, but it had been written for general practitioners. When the department faculty expressed interest in the findings on effort, it was agreed to conduct a single study designed to answer both lines of inquiry.

Faculty questions about the programmed text were as follows: Do medical and dental students actually learn what the program claims to teach?

Do the students find this program acceptable or enjoyable? What is the effect of not providing reinforcement after every instructional step or frame? The last question was raised by research which suggests that retention of learning might be improved as a result.

Traditional reinforcement theory (2) predicts that the most rapid learning occurs if every correct response is reinforced. It predicts that learning is slower under fixed ratio reinforcement, i. e., if every other, every third, or every tenth correct response, for example, is reinforced. The corollary prediction is that learning is slowest under variable ratio reinforcement, i. e., if there is no regularity with which the half, third, or tenth of all correct responses are reinforced.

Once a new behavior has been learned and reinforcement ceases, the theory predicts that it is then maintained longest by an individual who has learned it under a variable ratio schedule of reinforcement. This may be like the persistence of the slot machine addict who long ago won a jack-pot. Behavior is maintained only briefly by an individual who has learned under continuous reinforcement. Recall how few coins one inserts after a soft-drink dispensing machine breaks down. Behavior learned under a fixed ratio schedule is maintained for an intermediate number of trials.

That reinforcement theory did not predict the effect of effort on enjoyment and retention is of interest because the effect is predicted by dissonance theory.[1] Nor do the two theories agree upon the effect of fixed and variable ratio reinforcement. Dissonance theory predicts no difference between their effects on the maintenance of learned behavior, because the magnitude of dissonance is said to be determined by the absolute number of non-reinforced trials. If so, fixed and variable ratio reinforcement should not differ in the effect they produce on initial learning as well.

The two theories do have important areas of agreement. Both predict (a) that low effort produces more initial learning than high effort, and (b) that continuous reinforcement produces more initial learning than ratio reinforcement. Thus, the maximum initial learning achievement should be produced by a combination of low effort and continuous reinforcement. Both theories also predict (c) that behavior learned under ratio reinforcement is maintained longer after reinforcement ceases than behavior learned under continuous reinforcement.

---

[1] See Merrill and Buenz (1) for pertinent references and a brief discussion of Dissonance Theory.

It is now possible to state the points at which the two theories differ in their predictions regarding effort and reinforcement. The experimental purpose of this line of inquiry was to test the following hypotheses derived from dissonance theory:

I. Enjoyment of the learning situation is greatest under the conditions of high effort and minimum reinforcement (and least under low effort and continuous reinforcement). Reinforcement theory makes the opposite prediction.

II. Initial learning and retention are not affected by the difference between fixed and variable reinforcement at the same ratio. Reinforcement theory predicts that variable ratio will result in less learning and more retention than fixed ratio reinforcement.

III. Retention will be greater for learning produced under high effort than for learning produced under low effort. Reinforcement theory makes the opposite prediction.

Method

The study consisted of two experiments. Because of the intentional similarities in methodology, the main experiment is described first in detail, followed by a brief description of the important differences in the pilot experiment.

Main Experiment

The main experiment compared the effects of two levels of effort and four levels of reinforcement. The Low Effort level consisted of completing the responses called for in each frame of the programmed instruction, and the High Effort level consisted of completing all responses plus writing short essays justifying the responses made to 30 of the frames. This replicated and increased the effort conditions of the previous experiment (1). The 100% Reinforcement level used the programmed text in the normal manner, providing continuous reinforcement of every response; at 50% Fixed Ratio Reinforcement every other frame was reinforced; at 50% Variable Ratio Reinforcement half the frames were reinforced, but at random intervals; and at 0% Reinforcement no frame was directly reinforced. The 0% level did include indirect reinforcement from the redundancy of information within and among some of the frames themselves, and thus constitutes a variable reinforcement schedule at a ratio below 50%. Without the benefit of this indirect reinforcement no learning would be expected to occur.

The experimental sequence was the completion of the program, the immediate measurement of learning and attitudes, and after an interval of approximately a month, the measurement of retention.

Subjects. The 129 volunteer subjects (Ss) who completed the experiment included 72 second-year medical students and 57 second-year dental students who were enrolled in their respective courses in microbiology. These courses do not have identical sets of objectives, and because these and other differences might influence performance, the medical and dental students were given separate random assignments to one of the eight groups.

Materials. The section on the classification of allergies and anaphylaxis in the programmed text Allergy and Hypersensitivity (3) was used in the experiment. The different schedules of reinforcement were manipulated directly in the programmed text by having the appropriate reinforcers blocked out on the side panel of the text, which would normally reveal the correct response to each frame. Each S in the High Effort groups received a Supplementary Answer Sheet. It consisted of 16 pages 8-1/2" x 11" in size, and it indicated those 30 frames on which the short essays were required, allowing a space of half a page for each essay.

Measures. Members of the departmental faculty and staff prepared the 51-item objective test which covered the subject matter of the programmed instruction. The same items were used to measure learning and relative retention. A separate sheet contained two attitude questions which were used to measure perceived effort and perceived enjoyment.

Procedure. Prior to the first experimental session the students were invited to participate in a study " . . . to obtain an objective evaluation of the effectiveness of the programmed text and your personal comments based on this participation." They were also told that $2.00 would be donated to the class treasury for each student who completed the study. No mention was made of a second session at which retention would be measured.

During the first session the eight experimental groups met in separate but comparable classrooms. The session was held outside the regular class period because it lasted more than two hours. In each group the Ss first completed the specific form of the programmed instruction designated for their group. The High Effort groups also completed the Supplementary Answer Sheet while they were working through the program. As soon as all the Ss in each group finished their respective task they were asked to complete the achievement test, followed by the two attitude questions.

About a month later the second session was conducted without prior notification during a regular one-hour lecture period at which the entire class was present. To insure an independent response the Ss seated adjacent to each other received different forms of the test. Each of the three forms used contained a different random ordering of the 51 items. When

the test was completed, the full nature of the experiment was revealed, the results of the first session were reported, and the comments of the students were noted.

Because their class schedules were not the same, the dental students followed the procedure separately from the medical students.

Pilot Experiment

The pilot experiment was conducted approximately a year earlier than the main experiment. Its design anticipated the main experiment in that effort and reinforcement were varied for eight experimental groups. The sequence included the completion of the program and the immediate measurement of learning and attitudes only, although partial data on retention were obtained. A control group was first given the test of learning and then invited to work through the programmed instruction.

The 94 Ss in the pilot experiment were medical students in their second year, again anticipating the main experiment. Twenty-four Ss were placed in the control and 70 in the experimental groups on the basis of random assignment. The materials, measures, and procedure were the same as in the main experiment, except that no donation was made to the class treasury following the pilot experiment.

Results

All Ss complied with instructions, and in the pilot experiment the High Effort Ss wrote an average of about 720 words in completing the set of 30 essays. In the main experiment this average was about 740 words from the medical Ss and about 600 words from the dental Ss. This compares with an average of about 650 words from the nurses for the 20 essays required in the previous experiment (1).

Separate coefficients of learning-test reliability were prepared for each experiment and group of students. The learning test showed consistent reliabilities (4) in the pilot and main experiments of $r_{xx} = .57$ and $.63$, respectively. Encouragingly, the test was found to have a slightly higher reliability in the main experiment for the dental students than for the medical students, i.e., .64 vs. .50.

Computer facilities were used to ease the computational burden of the 15 analyses of variance required in the main experiment. Because the computer program was written to analyze equal frequencies within subclasses, Ss in the main experiment were deleted on a random basis until their number equalled 112, consisting of 56 medical and 56 dental students.

Considering the data on initial learning, for example, separate two-way analyses of variance (effort x reinforcement) were performed on the medical and dental student scores, followed by a three-way analysis (effort x reinforcement x student type) on those of all 112 Ss. These three analyses were also performed on retention, retention change, perceived effort, and perceived enjoyment scores. The analysis recommended for subclasses of unequal frequencies (4) was performed on the initial learning, the perceived effort, and the perceived enjoyment scores from the pilot experiment. There was only one statistically significant interaction effect in the total of 18 analyses of variance performed in this study.[2]

Tests of Hypotheses

Predictions in Common. Of those predictions common to both theories, predication (a), that High Effort produces less initial learning than Low Effort, was supported in the pilot experiment. The achievement of the High Effort groups was less than that of the Low Effort groups at every level of reinforcement, producing an $F = 6.36$ ($p < .05$). However, the means in the main experiment (Table 1) differed very little, and in all three analyses of variance the effect of effort was negligible. Prediction (a) was supported, but not consistently.

Prediction (b), that continuous (100%) reinforcement produces more learning than ratio reinforcement was strongly supported in the main experiment. In the case of the dental students, $F = 4.68$ ($p < .01$), the effect is attributable to the 0% level, a category of ratio reinforcement. On the other hand, the groups of medical students whichreceived 100% reinforcement scored higher than any of the ratio reinforcement groups, $F = 2.54$ ($p < .10$). When medical and dental students were combined for the three-way analysis of variance, the predicted effect of reinforcement produced an $F = 6.22$, far exceeding 4.04, the .01 level of statistical significance. In the pilot experiment, despite mean scores from 100% to 0% of 42.53, 42.06, 42.05, and 40.53, the $F = 1.50$ failed to approach statistical significance. Prediction (b) was strongly supported, but not consistently.

Prediction (c), that behavior learned under ratio reinforcement is maintained longer after reinforcement ceases than behavior learned under continuous reinforcement, was tested only in the main experiment. Retention change scores (initial learning minus retained learning scores) were prepared for each S. Smaller scores thus indicated that less forgetting had occured. The group means (Table 2) showed no consistent pattern upon inspection and negligible values of F in all three analyses of variance.

---

[2] Group means and other results of each analysis of variance are on file at the Communications Office for Research and Teaching, and individual copies may be obtained on request.

Prediction (c) was not supported.

TABLE 1

Initial Learning Means - - Main Experiment

Reinforcement		Effort: High	Effort: Low	Totals
0%	Med. Ss	37.00	37.71	37.36
	Dent. Ss	32.57	29.57	31.07
	All Ss	34.79	33.64	34.21
50% Variable Ratio	Med. Ss	38.43	39.57	39.00
	Dent. Ss	36.43	37.29	36.86
	All Ss	37.43	38.43	37.93
50% Fixed Ratio	Med. Ss	39.00	39.00	39.00
	Dent. Ss	36.00	36.00	36.00
	All Ss	37.50	37.50	37.50
100%	Med. Ss	40.71	41.43	41.07
	Dent. Ss	36.29	35.57	35.93
	All Ss	38.50	38.50	38.50
Totals	Med. Ss	38.79	39.43	39.11
	Dent. Ss	35.32	34.61	34.96
	All Ss	37.05	37.02	37.04

Antithetical Predictions. In order to test the three hypotheses unique to dissonance theory it was first necessary to assess the effectiveness of the manipulation of effort. It was necessary for the High Effort Ss to actually perceive that the experiment, including the set of 30 essays, was more effortful than did the Low Effort Ss. To answer the question How did you find this experiment, now that you have completed it?, the Ss checked their responses on an eight-point scale ranging from very effortful to not effortful at all.

The effort manipulation was consistently successful for the medical students in the pilot and main experiments, with F-ratios of 11.83 ($p<.01$) and 4.75 ($p<.05$), respectively. However, the High and Low Effort dental students did not differ in their perceptions, and the corresponding F-ratios were negligible. Thus the dental student scores were excluded from the tests of the antithetical predictions, but they provided aid in the interpretation of the medical student scores. The special case of the dental students

is considered in the Discussion. The analysis also considered the possibility that reinforcement under fixed and variable ratios might cause theexperiment to be perceived as more effortful than under the 100% reinforcement condition. This did not occur in either the pilot or main experiment for the medical Ss or for the dental Ss.

TABLE 2

Means for Retention Change[a]

Reinforcement		Effort: High	Effort: Low	Totals
0%	Med. Ss	4.29	4.86	4.58
	Dent. Ss	7.29	9.43	8.36
	All Ss	5.79	7.14	6.46
50% Variable Ratio	Med. Ss	8.57	3.00	5.79
	Dent. Ss	8.86	9.14	9.00
	All Ss	8.71	6.07	7.39
50% Fixed Ratio	Med. Ss	5.57	5.14	5.36
	Dent. Ss	6.57	8.14	7.36
	All Ss	6.07	6.64	6.36
100%	Med. Ss	9.00	6.57	7.79
	Dent. Ss	8.29	8.14	8.22
	All Ss	8.64	7.36	8.00
Totals	Med. Ss	6.86	4.89	5.88
	Dent. Ss	7.75	8.71	8.23
	All Ss	7.30	6.80	7.05

[a]Retention Change score = initial learning score - retained learning score, i. e., the smaller the mean the greater the retention.

The test of Hypothesis I was based on the question Now that you have the experiment, what did you think of it?. The Ss checked their responses on an eight-point scale ranging from enjoyed it very much (score 1) to did not enjoy it at all (score 8). Thus the groups with low mean scores indicated more enjoyment than groups with high mean scores.

High Effort groups did not perceive the pilot experiment as more enjoyable than did the Low Effort groups, $\underline{F} = 1.78$ (n. s.). This finding was repeated for the medical Ss in the main experiment, $\underline{F} = .02$ (n. s.). Groups receiving different levels of reinforcement did not differ in the perceived enjoyment of either the pilot or the main experiment, as indicated by the respective $\underline{F}$-ratios of .22 (n. s.) and 1.00 (n. s.). Thus this dissonance prediction was not supported in either experiment, nor was the antithetical prediction of reinforcement theory. Therefore, the test of Hypothesis I, that enjoyment of the learning situation is greatest under the conditions of high effort and minimum reinforcement, was inconclusive.

In another indication that the dental Ss constitute a special case, the Low Effort groups perceived the experiment as more enjoyable than did the High Effort groups, $\underline{F} = 4.84$ ($p < .05$). These data cannot be considered to support reinforcement theory, because the High Effort dental Ss did not perceive the experiment as any more effortful than did the Low Effort Ss.

Hypothesis II concerns the effect of fixed and variable ratio reinforcement. The effects they produced on initial learning were observed in the pilot experiment (no statistically significant difference) and in the main experiment. In the latter, differences were observed between ratio and continuous reinforcement, as reported for common prediction (b) but inspection (Table 1) establishes that 50% variable and 50% fixed ratio reinforcement produced no differences in initial learning achievement.

The effects on retention change (Table 2) were measured only in the main experiment, and here all $\underline{F}$-ratios were negligible. Hypothesis II, that initial learning and retention are not affected by the difference between fixed and variable reinforcement at the same ratio, was consistently supported.

Hypothesis III, the effect of effort on retention, was tested in the main experiment. Small differences in retention change score means were observed between High Effort and Low Effort Ss (Table 2), and the corresponding $\underline{F}$-ratios were all negligible. These findings failed to support either hypothesis. Therefore, the test of Hypothesis III, that retention is greater for learning produced under high effort than for learning produced under low effort, was inconclusive.

A statistical comparison of the performance of medical and dental Ss in an experiment based on medical subject matter is provided by the three-way analyses of variance. The medical Ss scored higher on initial learning, $\underline{F} = 28.80$ ($p < .001$); forgot at a slower rate, based on retention change scores, $\underline{F} = 8.77$ ($p < .01$); perceived the experiment as less effortful, $\underline{F} = 7.00$ ($p < .01$); and perceived it as similarly enjoyable, $\underline{F} = 2.93$ (n. s.).

Tests of Programmed Instruction

Do Students Learn? The pilot experiment was conducted toward the end of the course in microbiology. This provided a severe test of the programmed text, because the control group had profited from the conventional instruction and achieved a mean score of 30.0. Each of the eight experimental group means exceeded that of the control group, and the initial learning score mean for all 70 experimental Ss was 41.2. Recalling that a perfect score is 51, this difference represents a gain of more than 20% from the use of programmed instruction. Impressively, the students did learn.

The slightly lower mean initial learning score for all medical Ss in the main experiment, 39.11, may be accounted for by the fact that the main experiment was conducted early in the term, just prior to the start of conventional instruction in this subject matter.

TABLE 3

Mean Retention at One Month

Reinforcement		Effort: High	Effort: Low	Totals
0%	Med. Ss	32.71	32.86	32.79
	Dent. Ss	25.29	20.14	22.71
	All Ss	29.00	26.50	27.75
50% Variable Ratio	Med. Ss	29.86	36.51	33.21
	Dent. Ss	27.57	28.14	27.86
	All Ss	28.71	32.36	30.54
50% Fixed Ratio	Med. Ss	33.43	33.86	33.64
	Dent. Ss	29.43	27.86	28.64
	All Ss	31.43	30.86	31.14
100%	Med. Ss	31.71	34.86	33.29
	Dent. Ss	28.00	27.43	27.71
	All Ss	29.86	31.14	30.50
Totals	Med. Ss	31.93	34.54	33.23
	Dent. Ss	27.57	25.89	26.73
	All Ss	29.75	30.21	29.98

Is it Acceptable to Students? The attitude question on perceived enjoyment was used in both pilot and main experiments, and in each case the medical Ss rated their experience with programmed instruction between the mid-point of the scale and enjoyed it very much. However, during the discussion periods after both experiments the students made it clear that they did not want the programmed text used as a substitute for teaching-contact with the faculty. Since this clearly was not the intent, student attitudes toward use of the programmed text were considered favorable.

What Improves Retention? Evidence has already been presented which indicates the rate of forgetting, based upon retention change scores, did not differ among the four levels of reinforcement (or between levels of effort). Therefore, those groups which learned the most initially also remembered the most when tested a month later. Table 3 shows the main experiment mean scores for actual learning retained at the end of this interval. In the three-way analysis of variance, retention means were numerically smaller for the High Effort than the Low Effort groups, $F = .30$ (n. s.), and significantly smaller for the 0% Reinforcement level than for all others, $F = 3.18$ ($p < .05$). Interaction occurred between the effort and student variables, $F = 6.35$ ($p < .05$).

Discussion

After the previous experiment (1) it was asked if manipulating effort by requiring the writing of essays might not be a form of practicing the correct answers, which in turn might account for higher scores on the test of initial learning, and hence on the test of retention. The initial learning data in the present study confirm that this was not the case, because High Effort groups scored significantly lower in the pilot experiment and did not differ from the Low Effort groups in the main experiment.

The present study was designed to include an internal methodological check: if the predictions in common were confirmed, then greater confidence could be placed in findings favorable to the antithetical predictions of dissonance theory. The support of common predictions (a) and (b), which involved initial learning, indicates that the experiment actually manipulated effort and reinforcement in achieving the predicted results. Therefore, the lack of support of prediction (c), which involved retention, may challenge the validity of the assumption stated in the previous experiment (1), that resistance to extinction in animals (the maintenance of learned behavior after reinforcement ceases) is the equivalent of memory in humans. Both theories make their antithetical prediction about the effect of effort on resistance to extinction, and if human memory is not the equivalent, the findings on Hypothesis III might be expected to be inconclusive, as indeed they were.

If the retention assumption is not valid, Hypothesis II, concerning 50% fixed and variable ratio reinforcement, still supports the dissonance theory prediction about initial learning. Additional tests at ratios of other than 50% would provide broader support of this hypothesis.

The inconclusive findings on the enjoyment Hypothesis (I) are disappointing, because they suggest that the measure of perceived enjoyment may be unsatisfactory. If so, the possibility is increased that the statistically significant finding of more enjoyment among student nurses in the High Effort group in the previous experiment (1) is a Type II error which occurred by chance. The present measure replicated the one used in the previous experiment.

This study demonstrates the importance of selecting an appropriate programmed text, one which does not begin beyond the existing knowledge of the learners and which is written for a particular group. In such a situation, impressive gains in learning and relatively good student acceptance are the result. Where a text is inappropriate, as the text was found to be for dental students, its efficiency is reduced to a point at which continuous (100%) reinforcement produces no more initial learning than some type of 50% ratio reinforcement and where the perception of effort is significantly increased.

There is no reason to question the measure of retention, or the finding that students who learned more initially also remembered more than the others after approximately a month. The question of memory vs. resistance to extinction may affect the interpretation of the findings regarding the theoretical predictions involved. It does not affect the finding that manipulation of effort and reinforcement either interfered with student memory or failed to improve it. In this context, there is no reason for not using this programmed text as its authors intended.

III. THE EFFECTIVENESS OF MOTION PICTURES AT DIFFERENT STAGES OF LEARNING HISTORY-TAKING

Abstract

Eight film treatments demonstrating communication techniques in history-taking were presented at each of two stages of learning. The treatments differed according to four types of camera placement and an abstracted version (analogous to a line drawing) or a realistic version (analogous to a still photograph). At each stage the presentation to 88 medical students was followed by a test of learning achievement. Two attitudes were also measured: perceived learning and perceived enjoyment of the experiment. A separate study was conducted to determine if the least successful film treatment was effective. The results of the separate study indicated that viewing any of the film treatments improved learning achievement. The main study results supported the hypothesis that achievement is greatest if the abstract version is presented at the initial stage and the realistic version is presented at the later stages of instruction. The four camera angles had no differential effects on learning achievement. Student attitudes showed a high positive correlation between degrees of perceived learning and perceived enjoyment, but there was no correlation between degrees of perceived learning and actual learning.

* * *

Assume that one has the opportunity to produce a motion picture film which defines and demonstrates a number of communication techniques for effective patient history-taking. The film is to be shown to medical students during a course entitled "Introduction to Clinical Medicine." Previous investigations have suggested that the efficiency of such visual instruction is affected by the factors of camera placement, the degree of realism in the film, and the stage of student learning at the time the film is presented.

In an important early film study Roshal (5) compared the instructional efficiency of two angles of camera placement in teaching knot-tying, a simple perceptual motor task. For one film the camera was placed over the shoulder of the demonstrator at what Roshal termed the subjective angle because it enabled the viewer to see the task as if he were performing it himself. For another film the camera was placed opposite the demonstrator at what was termed the observer angle because the viewer faced the demonstrator while the latter performed the task. The subjective angle was found more efficient on a test of performance, and Roshal attributed this success to greater realism in the presentation, i. e., the subjective angle most closely approximated the actual task. In

a third film involving a rudimentary animation technique Roshal presented two ends of rope against a neutral background which then appeared to form themselves into the knot. This film was the most efficient of all three, but it did not fit either definition of cameral placement or support the hypothesis concerning realism.

In a review of these findings, Miller (6) noted that one of the problems of efficient teaching is to build up a discrimination so that the student will respond only to the relevant cues and not be misled or distracted by the irrelevant cues. The third film was based upon relevant cues, i. e., it was a kind of pictorial abstraction that starkly presented the features that distinguished the knot in question from any other. However, students also need realism in their training if they are to learn not to be misled by the large number of irrelevant cues which are present in many learning situations, and the first film (subjective angle) provided that realism. Miller suggested that one way out of this dilemma of relevant cue vs. realism is to use the various attention-focusing devices during the early stages of learning, and then after the student has learned to identify and respond to the relevant cues, gradually change the presentation in the direction of realism, eventually putting in all the distractions and variations that he is likely to encounter in the real-life situation. Thus, Miller called attention to an additional variable, stages of learning, and placed relevant cues against realism as dual aspects of effective visual information.

Television and film studies by Kumata (7) and Deutschmann, Barrow, and McMillan (8) investigated the initial stage of learning. They broadly defined the relevant cues as those that directly relate to the general objectives of instruction. In both studies there was an indication that use of a camera increased learning because the act of focussing on the subject matter tended to eliminate irrelevant details and distractions that may often be present in face-to-face instruction. Miller (6) states that further experimentation is necessary to clarify these issues.

Dietrich and Merrill (9) returned to the problem of camera placement in an experiment involving instructional material which is more complex than simple perceptual motor tasks such as knot-tying. They used kinescope recordings to teach hospital staff nurses the role of the team leader in the nursing care conference. The observer angle, which focused on the team leader only, produced more recognition learning that the subjective angle, which focused on the other four members of the nursing team. This superiority was retained for as long as four weeks. This finding was not similar to that of Roshal (5). Before the experiment the participating staff nurses were asked how much previous training they had received in this subject in order to classify them according to stage of learning. Based upon their replies, stage of learning was not related to the instructional

efficiency of either angle of camera placement. These findings could be criticized on two points: (a) four camera angles were possible in a situation involving two or more persons, and only two camera angles were tested, and (b) a judgment by the staff nurses was not the most accurate method for determining stage of learning.

Therefore, it was decided to replicate the Dietrich and Merrill (9) experiment and (a) include four possible camera angles in a situation involving only two persons, (b) vary relevance of cues and realism in terms of the objectives of instruction, (c) control the stages of learning during the course of the experiment, and (d) measure student perceptions of learning and enjoyment at both stages of learning. The specific hypotheses tested were as follows:

I. Learning about patient history-taking from films varies according to which of four camera angles is used to produce the visual information.

II. During the initial stages of learning, visual information is most effective if presented in an abstracted form, and during the later stages it is most effective if presented in detailed realism.

It was recognized that this experiment would involve preparation of a number of film materials of varying instructional efficiency. Therefore, it also was decided to conduct an independent study to see if all of the materials were effective.

Method

In the balanced experimental design eight film treatments of a patient history-taking session were varied at each of two stages of learning. The eight film treatments differed according to the four possible types of camera placement and two aspects of visual information. The four types of camera placement produced the subjective angle (showing patient only), the observer angle (showing physician only), the lateral angle (showing both physician and patient from a point at a right angle to their axis of eye contact), and the combined angle (showing both physician and patient full face simultaneously on a split screen). Different aspects of visual information were presented in a realistic version, a film treatment of an actual history-taking session, and an abstract version, prepared by emphasizing relevant cues within the realistic version and omitting detail considered irrelevant.

The experimental sequence was the presentation of the film at the initial stage of learning, the immediate measurement of learning and related attitudes, and after an interval of several months the presentation

of the same or different version of the film, followed by a repeated measurement. The camera angle seen at the initial stage was also seen at the later stage. Thus the complete experimental design called for 16 groups, 4 camera angles x 2 versions x 2 stages of learning (Table 2).

Subjects

A total of 88 volunteer second-year medical students completed the experiment. They were beginning their introduction to clinical medicine at the time of the initial film presentation. Course instruction continued during the interval preceding the later presentation, and this training included some practice in taking a patient history.

Materials

The title of the film material presented at each stage of the experiment was Communication Techniques in History-Taking. The film's viewing time was one hour, and it consisted of three segments. The first 24 minutes consisted of a lecture segment presented by two of the Experimenters (Es). It was the same for all Ss and briefly covered 40 teaching-points. The following 20 minutes contained the demonstration segment, in which a physician correctly illustrated the proper application of the teaching points with a patient during a history-taking session. This was the experimental segment of the film, and it was varied for each group according to camera angle and version. The lecture and demonstration segments together constituted the teaching portion of the film. The final 16 minutes consisted of the evaluation segment on which the written learning test was based. This segment was the same for all groups and is described under Measures.

The script for the realistic version of the demonstration segment was essentially a verbatim transcript of an actual history taken by one of the Es and recorded on videotape. It was selected because it illustrated the teaching points established as the instructional objective. The script for the abstract version was prepared by restating and emphasizing relevant cues in the realistic version that directly pertained to the teaching points and by omitting other material until the abstract and realistic scripts were of the same length. Subtitles were added to the abstract version for further emphasis of the relevant cues.

All four camera angles for a given version were filmed simultaneously with the same sound track. To avoid spurious differences between abstract and realistic versions, the roles of the patient and physician in both versions were played by the same two actors.

Measures

The 60-item objective test of learning was based upon the evaluation

segment of the film material. This segment was based on actual history taken by a senior medical student. It concerned a different patient and the roles were played by different actors. Half of the 40 teaching points were correctly demonstrated and the remaining half were incorrectly demonstrated. The written test measured how well the S had learned to recognize and label each point as presented in the film. An extra sheet contained questions which were used as a measure of perceived learning and perceived enjoyment. [1]

Procedure

Prior to the experiment the 88 Ss were told by members of the faculty that changes were being made in the introductory course in physical diagnosis and that some related film materials were being tested. Volunteer Ss were requested, and they were told that they would be reimbursed $10.00 for their participation in the experiment on two free afternoons. The groups met in eight classrooms on each occasion to complete the experimental sequence.

Independent Study

An independent study was conducted to measure the relative contribution of the lecture and demonstration segments to learning. Four groups were involved in the experiment. Group I viewed neither segment of the teaching portion of the film material, and it served as a control. Group II viewed the lecture segment; Group III viewed the demonstration segment; and Group IV viewed both lecture and demonstration segments. All groups then viewed the evaluation segment and took the same written test of learning used in the main experiment.

The experimental Ss in the independent study were 131 volunteer students at approximately the same level of training in a different medical school. They were placed in the four groups on the basis of random assignment, and they received no reimbursement.

Results

The coefficient of reliability (measuring internal consistency) of the 60-item test of learning as taken by the 88 Ss was $r_{xx} = .60$ (4). This compared with a reliability of .78 obtained earlier during a pilot test

---

[1] Film materials used in this experiment, along with copies of the related written test, are on file at the Communications Office for Research and Teaching.

involving the preceding class of medical students.

The Wilcoxen matched-pairs signed-ranks tests (10) were performed to test the effect of three procedural discrepancies upon learning. First, the scheduled three-month interval between stages of learning was not measured for all 88 Ss, because 65 Ss could not be reached until following their summer vacation. Second, about half the Ss experienced a film confusion at the beginning of the second presentation which might have affected their set to learn, even though it was corrected within a very few minutes. Third, 24 Ss completed the experiment believing that they would receive $2.00 rather than the $10.00 they actually received, a factor which might have affected their motivation to learn. The tests indicated that these factors had no statistically significant effect on the Ss learning scores, and the main analysis was completed as planned.

Learning

Two-way analyses of variance for unequal frequencies in subclasses (4) were performed on the data obtained from the initial test and the later test, and a third analysis considered the sequence of versions for the two stages of learning using combined data from both tests. Analysis of scores from a separate test indicated the effect of camera angle and abstract vs. realistic version at a known stage of learning.

Separate Tests. On the initial test the abstract version of the demonstration segment was superior to the realistic version, $F = 7.32$ ($p < .01$). The means (Table 1) for the abstract and realistic versions were 38.09 and 35.29, respectively. The first portion of Hypothesis II was supported. Analysis of the means produced by each camera angle indicated very similar results, $F = .66$ (n. s.). The actual means were Observer 37.68, Subjective 37.27, Lateral 35.71, and Combined 36.33. There was negligible interaction between the effects of version and camera angle, $F = .24$ (n. s.). Hypothesis I, on camera angles, was not supported in the initial test.

On the later test only a spurious comparison of the abstract and realistic versions was possible, since half the members in each of the eight groups had seen the opposite version on the initial test. As a result the difference between the spurious means was negligible, $F = .65$ (n. s.). However, each S viewed the same camera angle on both tests (although the versions may have differed), and this would have permitted the slight differences observed on the first test to increase on the second test. However, this did not occur; the numerical differences among the four means decreased, $F = .47$ (n. s.). There was negligible interaction between the effect of camera angle and (spurious) version, $F = .87$ (n. s.). Hypothesis I was not supported on the later test.

TABLE 1

Mean Learning on Initial Test

Camera Angle		Film Version: Abstract	Realistic	Total
Observer	(n)	(12)	(10)	(22)
	$\bar{X}$	39.50	35.50	37.68
Subjective	(n)	(11)	(11)	(22)
	$\bar{X}$	37.82	36.73	37.27
Lateral	(n)	(12)	(11)	(23)
	$\bar{X}$	37.50	33.82	35.74
Combined	(n)	(11)	(10)	(21)
	$\bar{X}$	37.45	35.10	36.33
Total	(n)	(46)	(42)	(88)
	$\bar{X}$	38.09	35.29	36.75

Combined Tests. The crucial analyses occurred when both sets of data were combined. The means in Table 2 were obtained after the scores on the initial and later test had been summed for each S. In considering Hypothesis II, each column represents one of the four possible combinations of the abstract and realistic versions with the two stages of learning. It can be determined from inspection that the highest mean score (78.04) was obtained from viewing both an abstract and a realistic version of the demonstration segment, provided the realistic version was presented at the later stage of learning. A comparison of the means obtained by presenting the same version at each stage indicates that the abstract version (76.59) was in general considerably more effective than the realistic version (69.19). The effects of the combinations were highly significant with an F-ratio of 6.32 with 3/72 d.f. ($p<.001$). These data provide strong support for Hypothesis II which states that during the initial stages of learning visual information is most effective if presented in abstracted form and during the later stages it is most effective if presented in complex detailed realism.

In considering Hypothesis I, the differences between the four camera angles in Table 2 were not significant, $F = .67$, as was the case in the separate tests. Even though the angles had a comparable effect, the observer and subjective angles were numerically superior to the combined and lateral angles. There was no statistically significant interaction between camera angle and combination of versions, $F = .99$. These data fail

to support Hypothesis I, which states that learning about patient history-taking from films varies according to which of the four camera angles is used.

TABLE 2
Mean Learning on Combined Tests

Camera Angle		Stage of Learning * / Film Version ** I A	II A	I A	II R	I R	II A	I R	II* R**	Total
Observer	(n)	(5)		(7)		(5)		(5)		(22)
	$\bar{X}$	81.60		78.57		75.40		68.60		76.27
Subjective	(n)	(6)		(5)		(6)		(5)		(22)
	$\bar{X}$	76.83		79.00		80.50		68.00		76.32
Lateral	(n)	(6)		(6)		(5)		(6)		(23)
	$\bar{X}$	72.00		80.67		73.20		69.17		73.78
Combined	(n)	(5)		(6)		(5)		(5)		(21)
	$\bar{X}$	76.80		74.00		73.80		71.00		73.90
Total	(n)	(22)		(24)		(21)		(21)		(88)
	$\bar{X}$	76.59		78.04		75.95		69.19		75.07

* Initial Stage (I), Later Stage (II)
** Abstract Version (A), Realistic Version (R)

Independent Study. The independent study was conducted using the film material from the main experiment that was found least effective at the initial test the material including the lateral camera angle for the realistic version of the demonstration segment. If this material produced a gain in learning, then all other materials used in the main experiment also could be considered effective. In the independent study, which was conducted at a different medical school, the group means for initial learning were as follows: I. control, 28.97; II. demonstration segment, 29.78; III. lecture segment, 31.91; IV. lecture and demonstration segments, 30.09. When the mean score of the 32 Ss in the control group was compared with the mean of all 99 Ss who viewed one or both segments of the teaching portion of the film (30.60), the expected difference favoring the latter was statistically significant, $t = 1.69$ ($p < .01$, one-tail). The difference between

Group III, lecture segment, and Group IV, lecture and demonstration segments, was not significant, $t = 1.59$ (n. s., two-tail).

Attitudes

To answer the question How much do you feel you learned from the experiment?, the 88 Ss in the main experiment checked their responses on an eight-point scale ranging from very much to very little. Data were collected at each stage tested. Analyses of variance performed on this measure of perceived learning were similar to those for actual learning. The groups did not differ in their perceptions of how much they had learned.

To answer the question Now that you have completed the experiment, how did you enjoy it?, the Ss checked their responses on an eight-point scale ranging from very much to very little. Here again, data were collected at each stage tested, analyses similar to those on actual learning were performed, and no statistically significant differences were found. The groups did not differ in their perceptions of how much they enjoyed the experiment.

When all 88 Ss were considered, a highly significant positive relation was found between perceived enjoyment and perceived learning ($X^2 = 18.95$, $p < .001$, during the initial stage, and $X^2 = 35.93$, $p < .001$, during the later stage). Those who enjoyed the experiment more also felt that they had learned more. The relation of each of these scales to actual learning scores was then explored. There was no relation between perceived enjoyment and actual learning (the Chi Squares were .72 and .17 for the initial and later stages, respectively). Nor was there any relation between what the students perceived they had learned and what they actually learned (the Chi Squares were .04 and .41 for the initial and later stages, respectively).

Discussion

Why the camera angles did not differ significantly is difficult to ascertain from the available data. There is a suggestion that the observer (physician only) and subjective (patient only) angles led to slightly greater learning achievement than the lateral and combined angles that focused on both physician and patient. This slight tendency suggests that the less complicated the visual information the greater the measured learning. The Dietrich and Merrill (9) findings on camera placement could be interpreted in this way, in that the visual information from the observer angle (nursing team leader only) could be considered less complicated than the subjective angle (all other nursing team members). The speculative nature of this suggestion underscores the shortage of experimental evidence available to teachers and other producers of instructional films regarding camera placement.

The highly significant findings related to stages of learning resulted solely from differences in visual information in the demonstration segment. All other aspects of the teaching portion of the film materials were identical for all Ss. The impressive effect produced by these differences upon measured learning is an indication of the potential importance of Hypothesis II. This is particularly true in the health sciences, where demonstrations of processes and techniques are an essential part of teaching. These findings suggest that motion picture film and television recordings of this nature will have far greater effectiveness if they are produced for use at a given stage of learning.

There is nothing in the rationale underlying Hypothesis II as provided by Miller (6) that indicates it is limited to instructional materials involving motion. It should also apply to the selection and preparation of charts, diagrams, graphs, and still photographs.

The separate study indicated that even the least efficient of the learning materials had some effectiveness, and it also indicated the importance of presenting visual information that is relevant to a given stage of learning. The separate study was conducted at the initial stage of learning, and the realistic version of the demonstration segment (appropriate to the later stage), when presented alone, contributed less than the lecture segment. When this demonstration segment was shown to Ss who had already viewed the lecture segment, it took 20 minutes of student time but contributed nothing to the achievement already produced. Reference to the group means suggests that the detailed realism of that demonstration even may have had some tendency to detract from the effectiveness of the lecture.

Because direct measurement of learning achievement is often difficult and time consuming, many studies of instructional presentations measure student attitudes instead of student learning. This approach is always plausible because of the tendency to believe that one has learned from instruction that was enjoyed, and the relationship was found in this experiment. The finding that actual learning had no relationship with either perceived learning or perceived enjoyment raises a serious question concerning the validity of attitude indicators as a measure of instructional effectiveness. It would appear that with present methods there is no valid alternative to direct measurement of learning achievement.

IV. TELEVISION RECORDINGS FOR PRACTICING PHYSICIANS:

THE EFFECT OF PROGRAM LENTH ON ATTENDANCE AND LEARNING

Abstract

The effects of massed versus distributed television presentations on attendance and learning in a voluntary situation were compared in a field experiment involving 114 physicians in general practice. The factorial design was based on four experimental groups, each located in a different community. Physicians viewed videotape presentations in a local hospital over a period of four weeks, with three one-hour sessions (distributed viewing) and a single three-hour session (massed viewing). On the fifth week they completed a learning achievement test over the televised instruction. The results indicate that attendance was significantly greater under massed than under distributed viewing, especially when the sessions took place in the evening rather than the morning. The presentations produced gains in learning achievement, and there was no statistically significant difference in learning achievement between massed and distributed viewing.

* * *

The first use of television to reach beyond the limits of the medical center for the continuing education of practicing physicians involved broadcasts in the late evening to physicians at home (11). This activity has progressively increased (12). The second pattern of usage also involved broadcast transmission, this time to groups of physicians meeting in the hospital around eight o'clock in the morning. The scrambled image broadcasts produced by Continuing Education in Medicine and the Health Sciences at the University of California San Francisco Medical Center and at the University of California at Los Angeles are representative of this second pattern.

In both patterns, the duration of the presentation is either approximately a half hour or a full hour in length. Duration, frequency, and time of day a presentation is transmitted are largely determined by pressures upon the schedule of the broadcast station, and no alternative can realistically be considered available. This is not the case for the third pattern, characterized by the operation of the Network for Continuing Medical Education and others, which involves noonday viewing of videotapes distributed by mail to groups in the hospital. However, Table 1 indicates that all patterns of television usage involve presentations of essentially similar duration.

The half hour and the one hour program may be most convenient for the broadcaster, but are they most desirable for the practicing physician? The recent availability of the dependable, low cost videotape recorder to

hospitals has turned this academic question into a practical one which is subject to measurement. The question has considerable importance, for it can affect all physicians who live beyond the coverage areas of broadcast stations employing the first and second usage patterns mentioned above.

TABLE 1

Television Usage Patterns for Continuing Medical Education in the United States

Distribution System	Viewing Situation	Viewing Location	Presentation Frequency	Presentation Duration	Time of Day
1. Broadcast	Individual	Home	1, 2, 4/mo.	1/2, 1 hour	10 p. m.
2. Broadcast	Group	Hospital	1/week	1/2, 1 hour	8 a. m.
3. Tape by mail	Group	Hospital	1/fortnight	1 hour	noon

This study asks if it is more desirable for physicians to view for a single three-hour session (massed viewing) or for three one-hour sessions at weekly intervals (distributed viewing). The popular short-courses offered to physicians on campus over a weekend resemble the massed viewing situation, and the excellent mail response from physicians at home to major continuing education symposia which are televised live in toto from the University of California Medical Center in San Francisco suggests that massed viewing is worthy of investigation.

The experimental evidence which is related to this question is for the most part indirect and confined to learning situations in the psychological laboratory. In the learning of simple tasks (e.g., tracing a pattern by mirror image, simple instrumental conditioning and rote memorization) distributed practice (periods of practice separated by periods of rest) is more efficient than massed practice (longer periods of practice with few or no interruptions). For example, an hour of practice divided into several shorter sessions will usually be more efficient in terms of greater learning than an hour of uninterrupted practice, provided each period of practice is long enough to allow at least one or more complete trials (2).

Findings arising from simple learning tasks may not be identical with findings based upon the highly complex subject matter of medicine. More importantly, the laboratory situation involving careful control of experimental

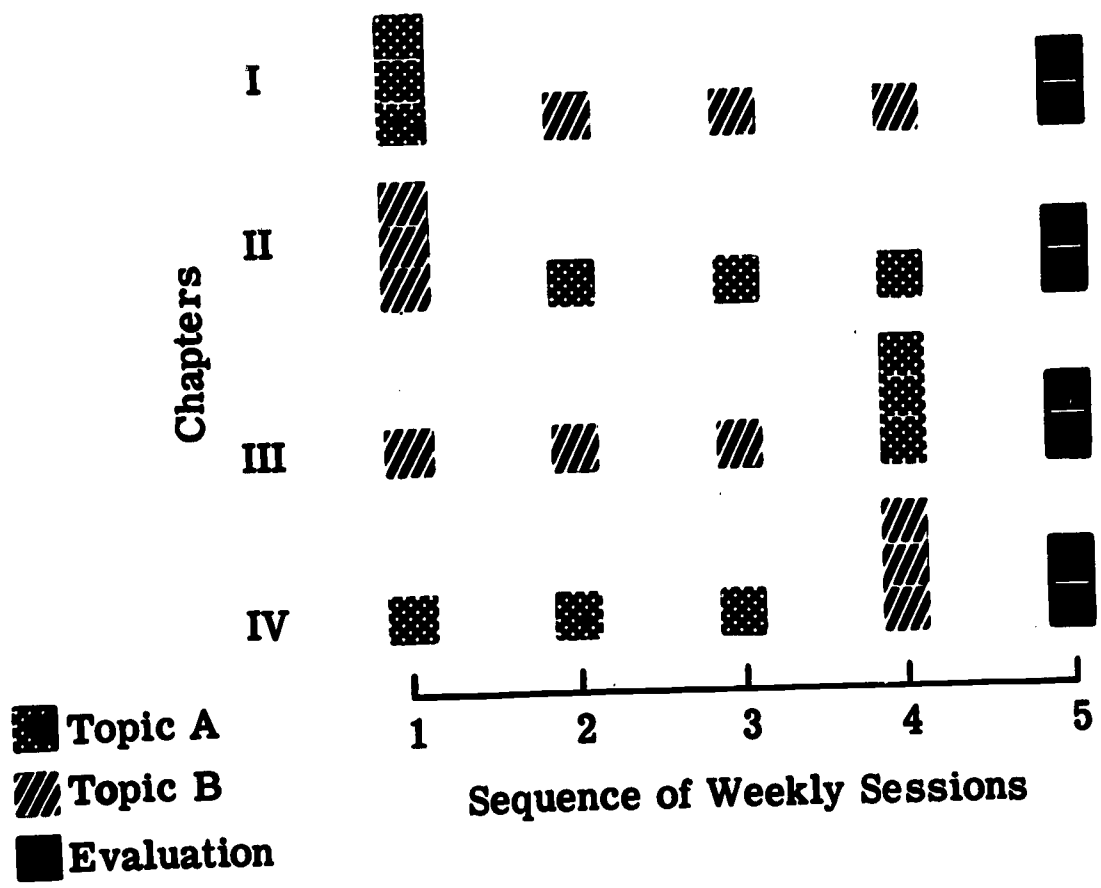
subjects (Ss) differs from the basic situation governing continuing education in medicine: e.g., practicing physicians constitute a voluntary audience inasmuch as attendance is a matter of personal choice.

If there is no attendance, learning from a television presentation is obviously impossible. Thus, it is important to learn how the length of a session affects attendance. However, attendance per se does not insure that the instructional objectives will be achieved, and the indirect experimental evidence suggests that the length of a presentation may affect the learning of those who attend. The efficiency of television presentations for a voluntary audience can be measured by observing the relation between attendance and learning, and the purpose of this experiment is to determine how massed and distributed viewing affect this relationship.

Method

The factorial design in Figure 1 indicates how two replications of the field experiment were conducted with the same practicing physicians. The replications differed only in the subject matter of the television presentations. Considering Topic A, it may be observed that the experiment extended over five weekly sessions, four being devoted to viewing and the fifth to direct testing of the presented subject matter. The design balances length of session (massed vs. distributed viewing) and order of presentation. The same applies to Topic B.

Figure 1. Experimental Design



Conditions were the same as those normally associated with this pattern of television usage; videotape sessions were held in hospital conference rooms, and attendance at all sessions was voluntary. Thus, the experimental observations could reflect the relative efficiency of the various presentations to the physicians in the "real world" of emergency calls, conflicting interests, and even physician fatigue. Each group was permitted to select the meeting time for the series, and three groups voted for 7:30 p. m., while one group voted for 8:00 a. m., all on a weekday.

A member of the research staff was present at every session to verify individual attendance. These data, plus the scores on the tests, constitute the measures of the two dependent variables. The optimum desirable length of presentation occurs if higher attendance is positively related with greater learning; the minimum, if lower attendance is related with less learning. With these two measures it is also possible to detect negative correlations, e. g., less attendance but greater learning for those who attend.

Subjects

The 114 Ss in the main experiment were practicing physicians from four communities judged demographically similar in northern California. Each of these Ss viewed one or more of the presentations described below. Physicians who may have indicated some intention of participating but who failed to appear were excluded from the analysis.

Each S was a member of his local chapter of the California Academy of General Practice. The instruction was sponsored by Continuing Education in Medicine and the Health Sciences of the University of California Medical Center in San Francisco. Participants were issued a Certificate of Attendance carrying one hour of credit for every hour of actual viewing. This is normal procedure whenever the University cooperates with the Academy. However, in this instance a registration fee was not charged because the physicians would be donating two hours of their time to participate in direct testing of what was learned from the videotape presentations.

Materials

Videotape Presentations. The six presentations, each of approximately one hour in length, were selected from the catalogue of the Association of Medical Television Broadcasters, using three criteria: Interest to general practitioners, quality of production, and independence of subject matter. They were divided into two topics, one of which included (1) "Clinical Diagnosis of Pulmonary Embolism," (2) "Chromosome

Studies--Why and When," and (3) "Acute Myocardial Infarction." The other topic included (1) "Cirrhosis Ward Rounds at Lemuel Shattuck Hospital," (2) "A Rational Approach to Low Back Pain," and (3) "Multiple Sclerosis." The criterion was independence of topics.

Along with a separate set of videotapes, each of the four hospitals was furnished identical sets of equipment, which included a portable videotape recorder and two 23-inch monitors.

Evaluation Test. A 120-item test of learning (with 20 items covering each of the six presentations) was used to evaluate the instructional effectiveness of the different experimental conditions and of the presentations themselves. The coefficient of reliability measuring internal consistency (4) was high (r_{xx} = .85 for 120 items; r_{xx} = .74 for Topic A; r_{xx} = .74 for Topic B). The items had been selected from 240 which were prepared for the pilot test taken by 24 volunteers from another chapter of the Academy which was located in a demographically similar area distant from the experimental locations.

Procedure

During the first four sessions of the experiment, the physicians viewed the presentations at the local hospitals under the conditions specified in the experimental design. Individual attendance was recorded at each session.

Prior to the first session, the physicians were informed (a) that the course was a cooperative venture between the University and the Academy to determine if the videotapes are an effective way for members to satisfy some of their study requirements and (b) that general practitioners are best qualified, and needed, to make such an evaluation. The longer viewing session was attributed to a limitation on the time the equipment was available.

During the evaluation session on the fifth week the physicians first took the 120-item learning test and then were invited to offer comment on the effectiveness of any aspect of the course. At the close, the full nature of the field experiment was revealed.

Results

Attendance

Examination of the data on all 114 participating physicians gives a general picture of the voluntary audience. Its essential characteristics may be observed by reading down the column headed Total in Table 2. The difference in the number of participants amoung the Groups reflects the difference in their total chapter memberships.

TABLE 2

Summary Data on Physicians' Voluntary Attendance to View Videotape Recordings

	Type of Participation		
	Viewing Only	Viewing and Evaluating	Total
	(n = 53)	(n = 61)	(n = 114)
Group		(Number of Physicians)	
I	16	18	34
II	18	14	32
III	7	12	19
IV	12	17	29
Total	53	61	114
Number of Presentations		(Percent of Physicians Viewing)	
6	17%	43%	31%
5	9	28	19
4	25	11	18
3	17	5	10
2	9	10	10
1	23	3	12
Total	100%	100%	100%
Sequence of Sessions		(Percent of Physicians Attending)	
1st	68%	74%	71%
2nd	51%	75%	64%
3rd	51%	85%	69%
4th	45%	80%	64%
Mean	54%	78%	67%
Topic	(Percent of Possible Physician-Hours of Attendance)		
A	53%	75%	65%
B	60%	85%	73%
Length			
Massed	62%	82%	73%
Distributed	51%	78%	65%
Total	57%	80%	69%

The amount of viewing was high. Of the 114 Ss who viewed at least one presentation, approximately a third (31%) viewed all six. Another third viewed either five or four presentations, while the remaining third viewed three or less.

If attendance were to begin at a high level and decline progressively after the initial session, then the presence of a novelty factor would be strongly suggested. Because topic and length were balanced over the four sessions, it is possible to inspect the data for such an order effect. Although a novelty factor cannot be ruled out, the fluctuation of ± 4% around a high mean attendance of 67% provides little or no evidence of its presence.

The percentages relating to number of presentations and sequence of sessions in Table 2 are computed from a divisor of 114. For example, 79 of the 114 Ss, or 69%, attended the third viewing session. A different divisor is indicated for computing percentages pertaining to topic, length, and all percentages which follow. For example, if each S in Group II (n = 32) attended two of the three possible distributed viewing sessions for Topic A, the Ss together account for 64 physician-hours, which when divided by 96 (32 x 3) represents an attendance figure of 67% of possible physician-hours of viewing.

With the balanced experimental design it was possible to determine that the attendance for Topic B (73%) exceeded that for Topic A (65%). Attendance for the entire experiment was 69%.

As for the major experimental variable, massed viewing (73%) proved superior to distributed viewing (65%). Attendance under distributed viewing was similar among all Groups (68%, 62%, 65%, and 68%). The massed viewing superiority was accounted for by Groups I, II, and IV (77%, 81%, and 76%) which viewed on weekday evenings. Group III had only 47% attendance when the three-hour viewing session was presented at 8:00 a. m.

More than half of the participants attended the testing session. The learning and attendance scores of these 61 Ss are used in the detailed analyses which follow. It is important to compare the 61 with the other 53 Ss (Table 2) to see if willingness to take an examination implies differences which would affect the generality of the findings. That those who had viewed more presentations would be more likely to take the learning achievement test is to be expected, and the difference between 57% and 80% in total attendance does not seem unduly large. In other respects the results are remarkably similar; Topic B was better attended, and massed viewing was superior.

Learning

Prior Knowledge. Although the Groups were selected on the basis of

demographic comparability, the data on learning (Table 3) indicate that the Groups differed in their prior knowledge of the subject matter. Groups III and IV scored consistently and significantly higher (Pair 2 mean = 60.90) than Groups I and II (Pair 1 mean = 47.7[illegible] despite the fact that attendance for Pair 1 (84%) was higher than that for Pair 2 (74%). All analyses to follow control for the difference in prior knowledge by performing separate statistical tests on each of the two Pairs.

TABLE 3

Mean Learning Scores and Attendance Percentages for Physicians Completing Evaluation Test

	Viewing Sessions							
	Massed			Distributed			Total	
	Topic	Learn. $\bar{X}$	Attend. %	Topic	Learn. $\bar{X}$	Attend. %	Learn. $\bar{X}$	Attend. %
Group								
I	A	23.22	78	B	24.56	81	47.78	80
II	B	24.29	93	A	23.36	88	47.64	90
(Pair 1)		(23.69)	(84)		(24.03)	(84)	(47.72)	(84)
III	A	28.00	58	B	29.00	67	57.00	63
IV	B	32.71	94	A	30.94	71	63.65	82
(Pair 2)		(30.76)	(79)		(31.14)	(69)	(60.90)	(74)
Topic A		25.13	70		27.52	78	26.35	74
Topic B		28.90	94		26.33	76	27.64	85
Total		27.05	82		26.93	77	53.98	80

Effectiveness of the Six Presentations. Did the six presentations teach; did the Ss learn from viewing? The mean score (score = number correct out of 20 items) of the Ss in Pair 1 who viewed was compared with that of the Ss not viewing, and so on, for each of the six presentations. This procedure was repeated for Ss in Pair 2, making in all a total of 12 comparisons. Eleven of these were positive, and in the one remaining instance the difference between means was only -.35. Application of the Sign Test (4) confirmed a statistically significant difference ($p < .01$), indicating the presentations were effective.

Order of Viewing. It is possible that the order of viewing might have influenced how much was learned. For example, under the condition of massed viewing did those who viewed Topic A in the first session learn more than those who viewed it in the fourth session, or vice versa? If the means for each Topic are rearranged by reference to Figure 1 to form a 2 x 2 table so that the columns reflect order and the rows reflect massed and distributed viewing, it can be seen from inspection that prior knowledge affects the column totals. When columns are summed over both Topics, the effect of prior knowledge on order is balanced out, and the sums of the means are then seen to differ only slightly (Viewed First = 57.45; Viewed Second = 58.53). The probability of an order effect is accordingly rejected.

Length. The Topic A learning scores of each S in Group I (massed viewing) were compared with those of each S in Group II (distributed viewing) by means of the Wilcoxen Two-Sample Test, or Mann-Whitney Test (10) in order to produce the z score for Pair 1 (Table 4). This procedure was repeated for Ss in Groups III and IV, and for all additional learning comparisons shown. There were more comparisons favorable to distributed (-) viewing in each instance with the major exception of Pair 2 under Topic B, but in none of the instances did these differences approach $p = .10$. Length was not shown to have any statistically significant effect on the amount learned by a voluntary viewing audience of practicing physicians.

Following a similar procedure, the Topic A attendance of each S in Group I was compared with that of each S in Group II, and so on, by means of Armitage's (13) or Terpstra's (14) Trend Test. When all four Groups were combined, attendance under massed (+) viewing was found significantly superior to that under distributed viewing ($p < .01$). This was also true for massed viewing in theseparate analysis of Topic B. In the analysis of Topic A, massed and distributed viewing were not shown to differ from each other. These data indicate that under the conditions of this field experiment length has a statistically significant effect on attendance, which is greater for massed viewing.

Discussion

There is some indication that the well-established superiority of distributed practice under laboratory conditions is operating within the data on the voluntary audience of practicing physicians. Where attendance is equal, as for Topic A, distributed viewing shows a tendency to produce greater learning. Even where massed attendance is very much higher, as for Topic B, the gains in learning are only reduced to approaching equality.

TABLE 4

Superiority of Massed (+) vs. Distributed (-) Viewing for Attendance and Learning by Topic

		Attendance		Learning	
		Superiority	z-Score[a]	Superiority	z-score[b]
Topic A					
Pair 1			0.		- .38
Pair 2			-.02		-1.44
(Pairs 1 and 2)			(-.02)		(-1.23)
Topic B					
Pair 1	Massed		1.88*		- .95
Pair 2	Massed		3.55***		1.02
(Pairs 1 and 2)	(Massed)		(3.81)***		(- .06)
Both Topics					
Pairs 1 and 2	Massed		2.60**		- .91

[a] Terpstra's Trend Test

[b] Wilcoxen Two-Sample Text

* p < .10
** p < .01
*** p < .001

However, the z scores (Table 4) clearly indicate that since these apparently consistent differences do not approach one of the standard levels of statistical significance there is no basis for assuming that they are determined by anything other than chance. Table 3 also illustrates the similarity in amount of learning gained. In fact, the mean learning score under massed viewing, 27.05, is greater than that under distributed viewing by .12. The most reasonable conclusion to be drawn from the findings is that massed viewing attracts greater attendance than distributed viewing with no significant reduction in learning achievement.

Attention is called to the time of day selected by the four Groups for viewing. These data may be of interest to those who are planning to offer

a continuing education course by videotape. The understandable reluctance of physicians to devote a whole weekday morning to massed viewing makes this finding particularly applicable to weekday evening viewing.

There is a practical implication to be drawn from the large difference in the attendance for Topic A and Topic B. It applies equally to those who viewed and to those who also participated in the test. Practicing physicians do not voluntarily attend such sessions for the novelty of viewing videotape recordings. The subject matter affects attendance.

Neither the Experimenters nor the Officers of the California Academy of General Practice could predict the extent of the cooperation to be expected in a field experiment of this nature. The participation of a total of 138 practicing physicians (including those in the pilot test), the high rate of attendance (69%) in the main study, and the participation of 61 in the testing session should be particularly noted. This is evidence that physicians will voluntarily participate in research on the teaching effectiveness of a medium of communication if the relevant nature of the study is explained in advance and if the need for their cooperation is made clear.

APPENDIX

ADVISORY COMMITTEE AND RESEARCH STAFF

The Committee

Curtis, (Mrs.) Orson L., R.N., M.A., (Frieda L. Smith) formerly Associate Professor, College of Nursing, University of Arizona.
Dryer, B. V. (M.D.), Medical Director, Media Medica, Inc., New York.
English, J. A. (D.D.S.), Dean, School of Dentistry, State University of New York at Buffalo.
Ramey, J. W. (Ed. D.), Executive Director, Institute for Advancement of Medical Communication, Philadelphia, Pennsylvania.
Smoke, Mary E. (Ph.D.), Associate Professor, Department of Mathematics, California State College, Long Beach.

The Staff

Principal Investigator
Merrill, I. R. (Ph.D.), Director, Communications Office for Research and Teaching. Served May 1, 1965, to August 31, 1967, and January 1 to June 30, 1968.
Yaryan, Ruby B. (Ph.D.), Program Coordinator, May 1 to 31, 1965. Project Psychologist, May 23, 1966, to August 31, 1967. Acting Principal Investigator, September 1 to December 31, 1967.

Co-Investigator
Farber, S. M. (M.D.), Dean of Educational Services and Director of Continuing Education in Health Sciences.

Television Technical Advisor
Litke, R.A., Senior Development Engineer, Communications Office for Research and Teaching.

Program Coordinator
Musser, T. S. (M.A.), August 30, 1965, to December 31, 1967.

Medical Script Writer
Bell, Louis, January 3 to February 10, 1966.

Principal Television Technician
Mudge, Richard, May 1, 1965, to May 2, 1966.

Administrative Assistant
Ward, T. L., September 5 to December 31, 1967.

Project Secretary
Sanchez, Anna J., May 13, 1965, to January 2, 1967.
Dwyer, Katherine, January 3, 1967, to September 29, 1967.
Allen, Dorothy, November 13 to December 31, 1967.

REFERENCES

1. Merrill, I. R., and Buenz, Ruby Yaryan. Retention and Enjoyment of Medical Training: the Effect of Effort. J. Med. Educ., 41: 230-236, 1966.

2. Berelson, B., and Steiner, G. A. Human Behavior: an Inventory of Scientific Findings. New York: Harcourt, Brace & World, 1964.

3. Basic Systems, Inc. Allergy and Hypersensitivity. (2nd ed.) New York: Chas. Pfizer & Co., Inc., 1964.

4. Walker, Helen M., and Lev, J. Statistical Inference. New York: Holt, 1953.

5. Roshal, S. M. Effects of Learner Representation in Film-Mediated Perceptual Motor Learning. Technical Report SDC 269-7-5. Port Washington, N. Y.: Special Devices Center.

6. Miller, N. E. Graphic Communication and the Crisis in Education. (Special issue AV Communication Rev., 5, 3, 1957.) Washington, D.C.: Dept. of Audio-visual Instruction, National Education Assoc., 1957.

7. Kumata, H. Two Studies in Classroom Teaching by Television. In The Impact of Educational Television, Schramm, W. (Ed.). Urbana: Univer. Illinois Press, 1960.

8. Deutschmann, P. J., Barrow, L. C., Jr., and McMillan, Anita. The Efficiency of Different Modes of Communication. AV Communication Rev., 9: 263-270, 1961.

9. Dietrich, Betty J., and Merrill, I. R. Television in Health Sciences Education: Camera Placement. Nursing Res., 13: 217-221, 1964.

10. Siegel, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.

11. Castle, C. H. Open Circuit Television in Postgraduate Medical Education. J. Med. Educ., 38: 254-260, 1963.

12. Merrill, I. R. Open Circuit Television in Continuing Education in the Health Professions, NAEB J., 26: 4, 15-21, 1967.

13. Armitage, P. Tests for Linear Trends in Proportions and Frequencies. Biometrics, 42: 375-386, 1935.

14. Terpstra, T. J. The Asymptotic Normality and Consistency of Kendall's Test Against Trend, When Ties are Present in One Ranking. Indagationes Mathematicae, 14: 327-333, 1952.